

# Dealing with Fake News: Policy and Technical Measures

*On behalf of the Massachusetts Institute of Technology (MIT) Internet Policy Research Initiative (IPRI)*

Fake news has, by many definitions, existed throughout human history, from early word of mouth transmission of news, to the printing press which allowed for greater dissemination of ideas, and to the internet and social media services that provide a convenient and cheap means of dissipating news globally. While social media has democratized content creation, provided a platform for previously overlooked local stories and experiences, offered a means of expression for those living in closed societies, spurred small businesses as well as allowed for political organization and activism, it has also introduced new tools for actors who seek to gain from the proliferation of fake news. In recent times, the prevalence of fake news on social media platforms has occurred in close proximity to democratic elections in various countries. Although determining the exact influence of fake news on election outcomes amid a wide range of other factors remains an open question, it is important to address the complex socio-technical challenges posed by the prevalence of fake news on social media platforms.

Apart from the high speed of online news transmission and the low barrier to its creation across borders, several other factors have contributed to making fake news significantly more fast-paced and technically sophisticated at present. First, the use of algorithms to quickly sift through massive amounts of data and identify relevant sub-populations of people for highly personalized advertisements also allows for pernicious targeting<sup>1</sup> or micro-targeted manipulation, which aims to trigger emotive responses to persuade particular audiences. This also has the effect of furthering polarization among groups of people by increasing both the prevalence and isolation of echo-chambers, where individuals only receive information that coincides with their pre-existing beliefs and perspectives. Second, the engagement-driven advertising business model of popular social media platforms tends to promote controversial or provocative articles, often ignoring the veracity of the content while ensuring high visibility and seamless integration of fake news alongside trusted content. Third, the prevalence of bots and human-administered fake accounts (or “sock puppet accounts”) on social media platforms enables malicious actors to hide or misrepresent their identities, amplify messages to push a particular agenda, and drown out oppositional voices. When combined, these three factors allow for unprecedented sophistication in reaching large audiences and misleading online news consumers.

The existence of fake news on platforms that have become central to the everyday lives of citizens continues to exacerbate political polarization and social mistrust, undermining fair participation in online discourse and creating a subtle but significant risk of future voter suppression and election interference by foreign and domestic actors. If not appropriately tackled, the propagation of fake news will make it increasingly difficult for democratic states to advance their citizens’ interests and

---

<sup>1</sup> For instance, ProPublica revealed that Facebook allowed advertisers to target algorithmically-identified “Jew haters.” See: Angwin, Varner, and Tobin. “Facebook Enabled Advertisers to Reach ‘Jew Haters’ — ProPublica.” September 14, 2017. Accessed February 27, 2018. <https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters>.

resolve social issues through public deliberation. New interventions are needed to incentivize social media platforms to make necessary changes to remove existing vulnerabilities that undermine democratic discourse and principles.

We explore three broad issues at the heart of the fake news phenomenon: increasing microtargeting in political advertising, the departure from editorial standards in the broadcasting of news on social media and the potential for spreading disinformation at scale through means such as bot accounts impersonating humans, inauthentic accounts, and fake video content. In tackling each of these issues, we suggest areas for potential policy and technical interventions. To compensate for the information asymmetry in microtargeting between advertisers and users of social media platforms, we recommend mandated disclosures for political campaign advertisers, disclosure requirements for all targeted advertisements and third-party applications gathering data for use in microtargeting, and platform design features that offer more diverse political ads and greater choice in viewing ads to users. In order to compensate for social media platforms' departure from editorial standards in news broadcasting, we suggest creating incentives that reward social media publishers based on content quality, accuracy or depth instead of only engagement metrics, and offering tools for greater user control of their personalized news experiences. For tackling impersonation and disinformation at scale, we explore the possibilities of labelling bot-operated accounts, authenticating influential accounts and pages, and verifying fake video content.

## Key issues and solution areas

### 1. Political advertising: increasing microtargeting

While the past few decades have seen various efforts to leverage digitally-mediated commercial advertising for political campaigning purposes, the regulation of political advertising remains a complex and charged issue. The ability of citizens to express their ideas and participate in political discourse is fundamental to democracy, and must remain protected by rights to freedom of expression and to political participation on all platforms.<sup>2</sup> However, unfettered access to unregulated paid advertising on social media platforms not only provides domestic and foreign actors with a medium for publishing fake news but also offers tools to actively promote dissemination.

In the United States, the most visible and well-known direct regulation of the use of broadcast media by political advertisers is the “equal time rule”, which requires radio and television stations and cable networks to treat legally qualified political candidates equally in allocating airtime.<sup>3</sup> While this ensures that candidates receive the opportunity to purchase the same amount of airtime with the same audience size at the same rate, this rule does not apply to the print media or to political advertising by third parties, which is almost entirely unregulated. The obligation of

---

<sup>2</sup> Universal Declaration of Human Rights, UN General Assembly Resolution 217A(III), 10 December 1948, Articles 19 and 21.

<sup>3</sup> Communications Act of 1934, 47 U.S.C. § 307(a).

neutrality in broadcast media is one of the rare features common among various regulatory systems. In contrast to the minimalist regulatory approach taken in the United States, several European states, which also tend to be nations with a strong tradition of public broadcasting, prohibit paid campaign advertising on radio and television altogether.<sup>4</sup> In the United Kingdom, for instance, all paid political advertising is completely prohibited on television and radio; instead political parties are given airtime via party political broadcasts, which are not classified as advertising. Apart from political candidates and parties, this prohibition applies to any advertisement aiming to influence public opinion on a matter of public controversy. Meanwhile, political ads in non-broadcast media (posters, newspapers etc.) whose principal function is to influence voters in local, regional, national or international elections or referendums are exempt from the Advertising Code, and cannot be investigated by the Advertising Standards Authority (ASA), the U.K. regulator of advertising. While paid political advertising is permitted in newspapers and on billboards, strict restrictions also exist on printing and publishing by third parties during political campaigns in the U.K.<sup>5</sup>

Given this context, there has been considerable development in political campaigning since the earliest broadcasted ads and the earliest candidate websites, when there was no online advertising industry specializing in politics and campaigns were generally limited to running banner advertising across the Web. With technological innovations in online ad targeting, campaigns have moved away from tethering ads to geographical units to reach broad publics towards designing content to influence the vote choices of particular groups and individuals. Through cookie-enabled web-tracking and email tracking, political campaigns have been able to track which websites a user visits and whether they opened a particular campaign email, which is then used to personalize messaging to each potential voter through not only direct mailing, but also through emails and ads across the Web dealing with specific issues. In 2008, Facebook began to allow political firms access to users' profile data to help facilitate targeted advertisements on its platform and Twitter also began seeing substantial political use. The use of social media platforms to discern political affiliations and positions on a wide spectrum of socio-political issues makes them a prime enabler of using targeted ads to spread disinformation and also generate revenue while doing so.

While campaigns have long used different mediums, such as radio, broadcast and cable television, to deliver select messages to groups based on geographic location and audience demographics (such as profiles of cable channel viewing audiences), what differentiates micro-targeted ads from their mass broadcast counterparts is the greater ease and often lower cost of operation as well as the greater accuracy of the data that lies behind them: advertisers can more finely target various audiences by examining the online behaviors, lifestyles, likes or dislikes of known audiences and then seeking out "lookalike audiences". Instead of broadcasting a consistent message to a diverse audience (historically across many different media), micro-targeting on social media platforms allows an actor to direct different messages to narrow subsets of users, thereby reinforcing and amplifying partisan, group, and identity conflicts. This can also allow advertisers to use data about

---

<sup>4</sup> "Regulation of Paid Political Advertising: A Survey." Center for Law and Democracy. March 2012. Available at: <http://www.law-democracy.org/wp-content/uploads/2012/03/Elections-and-Broadcasting-Final.pdf>.

<sup>5</sup> Id.

in-group biases, bandwagon effects, or psychological state to identify and mobilize individual users or groups with particular messaging designed to incite them.

The increase in the amount of personal data available to social media advertisers is likely to continue unabated as individuals increase their online activity. The risk remains that highly targeted political and issue-based ads that spread disinformation will limit democratic debate, lead to increasingly selective exposure and polarization, erode individuals' privacy and affect the competitiveness of elections. Some possible measures to address these issues are listed below.

### 1.1 Mandating disclosures for political campaign advertising on online platforms

Since social media platforms allow advertisers to reach both a wider audience and a far more narrowly defined audience than conventional broadcasting, ads on social media platforms from political campaigns should be required to carry the same disclaimers from their sponsors as do TV, radio and print ads, namely explicitly stating who paid for and approved the ad. Currently, in the United States, according to the Federal Election Commission's rules of advertising, any public communication made by a political committee is required to have a disclaimer that states whether or not the communication was authorized or financed by the campaign or candidate.<sup>6</sup> Similarly, under California's DISCLOSE Act, all electronic advertising usually must include sponsorship information, in the form of a link that reads, "Who funded this ad?," a direct link to sponsor information, or include the sponsor information in the ad itself.<sup>7</sup> In the UK, candidates and non-party campaigners are required to include an "imprint," or disclaimer, on physical advertising that must include the name and address of the printer, the promoter who caused it to be published, and the person on whose behalf the material is being published.<sup>8</sup> In 2011, Facebook requested the U.S. Federal Election Commission (FEC) to exempt it from rules requiring political advertisers to disclose who is paying for an ad a year after the FEC granted such an exemption to Google.<sup>9</sup> Extending such disclosure requirements to social media platforms would help ensure that social media users can easily distinguish between official campaign-sponsored ads and all other third-party ads.

#### Implementation

At minimum, political campaign advertisements on social media platforms should have the same transparency and disclosure requirements imposed as for conventional broadcast and print media. Platforms should also be required to disclose the full extent of their advertising practices, and

---

<sup>6</sup> "Advertising." FEC.gov. Accessed February 24, 2018.

<https://www.fec.gov/help-candidates-and-committees/making-disbursements/advertising/>.

<sup>7</sup> "The California DISCLOSE Act: California Clean Money Campaign." Accessed February 24, 2018.

<http://www.caclean.org/progress/ab1148.php>.

<sup>8</sup> Representation of the People Act 1983, s 110(2)(a), (3) to (6) and (13).

<sup>9</sup> Neidig, Harper. "Facebook Fought for Exception to Political Ad Disclosure Laws in 2011." Text. The Hill, September 28, 2017.

<http://thehill.com/policy/technology/352848-facebook-fought-for-exception-to-political-ad-disclosure-laws-in-2011>.

whether or not political candidates get the same rates for placing similar ads as well as the amounts paid for political ads.

## 1.2 Imposing disclosure requirements for all targeted advertisements and third-party applications gathering data for use in microtargeting

To hold users' attention, platforms use specifically tailored ads and posts. This mechanism predominantly helps advertisers who can use the highly personalized data (past clicks, 'likes', shares, etc.) to more effectively target audiences. However, the continuous reinforcement of existing ideas in this manner can facilitate the establishment of beliefs that are more deeply entrenched, more extreme and more resistant to contrary facts. Such microtargeting can also facilitate manipulation by political advertisers, who create numerous ad variations, and then narrowly target them towards various niche groups, thereby limiting access to accurate information among like-minded communities and promoting polarization on a grander scale. The micro-targeted messages allow advertisers to control the timing, information, and sites of contact, so that they leverage psychological predispositions or vulnerabilities for maximum effect. Moreover, because micro-targeted posts are typically only viewed by targeted audiences, there are limited opportunities for critique, correction or counterargument.

Since fake news is often propagated using ads focused on polarizing issues to amplify existing social divides and partisan conflict, there is a need to extend advertising disclosure provisions to all issues-based advertising in addition to ads for political candidates or campaigns. For instance, during the 2016 U.S. presidential election campaign, Russian operatives bought ads from Facebook targeted at "professed gun lovers, fans of Martin Luther King Jr., supporters of Trump, supporters of Clinton, residents of specific states, and Southerners who Facebook's algorithms concluded were interested in 'Dixie.'"<sup>10</sup> Such issue-based ads may not particularly mention political candidates or elections. Even if the definition of a political ad is radically expanded and the expenditure of all political ads is carefully tracked, any advertiser can still easily make their own fake news website then pay platforms to boost their "news" articles to a broader audience. More generally, the ease of hiding the location or identity of authors of social media posts or bots makes election regulations on campaign spending and media broadcasts extremely difficult.

In light of these factors, beyond indicating the source of an ad, there should be disclosure requirements for all targeted ads, which allow for more fine-grained targeting than that possible via other means. The Honest Ads Act introduced recently in the U.S. Senate contains several requirements, including an obligation for online platforms to provide a public record with, "a description of the audience targeted by the advertisement, the number of views generated from the advertisement, and the date and time that the advertisement is first displayed or last displayed."<sup>11</sup>

---

<sup>10</sup> Craig Timberg, Elizabeth Dwoskin, Adam Entous and Karoun Demirjian, "Russian ads, now publicly released, show sophistication of influence campaign", Washington Post (Nov. 1, 2017), [https://www.washingtonpost.com/business/technology/russian-ads-now-publicly-released-showsophistication-of-influence-campaign/2017/11/01/d26aead2-bf1b-11e7-8444-a0d4f04b89eb\\_story.html?utm\\_term=.20b50dbb6554](https://www.washingtonpost.com/business/technology/russian-ads-now-publicly-released-showsophistication-of-influence-campaign/2017/11/01/d26aead2-bf1b-11e7-8444-a0d4f04b89eb_story.html?utm_term=.20b50dbb6554).

<sup>11</sup> Honest Ads Act, S. 1989, 115th Cong. (2017-18), SEC 8 (a)(2)(B).

While Facebook has introduced transparency mechanisms such as the “Why am I seeing this?” button to offer ad explanations and an Ad Preferences Page that offers data explanations to each user, recent studies have found that “ad explanations are often incomplete and sometimes misleading while data explanations are often incomplete and vague”<sup>12</sup> and that users often found the vague and oversimplified language of ad explanations uninterpretable and sometimes untrustworthy.<sup>13</sup>

Secondly, in addition to advertising disclosures, there should also be greater visibility and transparency of the access permissions used by third-party applications on social media platforms to gather data from users for more accurate profiling and microtargeting — including, potentially, explicit purpose limitations for the derivative or third-party use of that data. Given the recent news of how Cambridge Analytica, a UK-based political consulting firm, had inappropriately harvested data from the Facebook profiles of over 50 million people, most of whom did not provide consent for their data to be used for political and psychological profiling,<sup>14</sup> the need for social media users to be able to better view, understand and manage access permissions for applications that collect their personal data for use in microtargeting is evident.

### Implementation

Apart from explicitly indicating the entity placing the ad, disclosures for targeted ads must unambiguously explain the criteria used to target an audience and the data collected about the user that caused the user to be placed within that target audience in order to provide correct and complete explanations for why certain users were targeted with particular ads.

Past research has demonstrated that when users can see what personal information is being sent by the smartphone applications they use to which third parties, application users are to make more reasoned choices about data sharing and application usage.<sup>15</sup> In a similar fashion, more transparent and explicit disclosures of all the access permissions available to various third-party applications on social media platforms could help increase awareness among social media users about the applications that are collecting their personal data for use in posting micro-targeted ads to them. This would also enable social media users to make more informed choices regarding which third-party applications to use on social media platforms.

---

<sup>12</sup> Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna P Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook’s Explanations. In *The Network and Distributed System Security Symposium (NDSS)*.

<sup>13</sup> Eslami, M., Kumaran, S. R. K., Sandvig, C., Karahalios, K. (in press). Communicating Algorithmic Process in Online Behavioral Advertising. *Proceedings of the 36rd Annual SIGCHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery (ACM).

<http://web.engr.illinois.edu/~eslamim2/publications/eslami-CHI18-ads.pdf>

<sup>14</sup> Graham-Harrison, Emma, and Carole Cadwalladr. “Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach.” *the Guardian*, March 17, 2018.

<http://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.

<sup>15</sup> Van Kleek, Max, Ilaria Liccardi, Reuben Binns, Jun Zhao, Daniel J. Weitzner, and Nigel Shadbolt. “Better the Devil You Know: Exposing the Data Sharing Practices of Smartphone Apps.” In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 5208–5220. CHI ’17. New York, NY, USA: ACM, 2017.

<https://doi.org/10.1145/3025453.3025556>.

The disclosures for both targeted ads and the access permissions available to third-party applications to collect personal data for microtargeting on social media platforms should be clearly visible and interpretable by social media users to help them understand how their data is being used and also enable advertisers to offer more credible and trustworthy ads.

### 1.3 Maintaining the essence of the ‘equal time’ principle in online political campaign advertising and other platform design features

Outside of social media platforms, political candidates buy attention in the form of airtime, newspaper ad space, or direct mail postage at roughly similar rates to reach the same amount of people. The 1927 U.S. broadcasting equal-time rule<sup>16</sup> obligated radio and television stations to offer equivalent time at the same price to political candidates. One way of maintaining the essence of this rule on social media platforms is to create different standards for official campaign-sponsored political ads as opposed to commercial or third-party ads.

#### Implementation

A possible means of achieving this on social media platforms would be to reduce reliance on engagement metrics in the platforms’ internal ad auctions for official campaign-sponsored political ads and recommend ads posted by all other political candidates who have paid equally to place their ads on the platform to users interacting with official campaign-sponsored political ads. This would result in richer and more comprehensive information in news feeds in an attempt to encourage multifaceted and holistic discussions, offer different viewpoints and remove siloes or echo chambers as it pertains to election-related issues.

Additionally, design tools could be introduced that allow users to opt out of receiving targeted ads in their social media feeds based on certain specified data categories (similar to the data categories available to advertisers for posting targeted ads). Apart from the disclosure requirements discussed in 1.2, another option that would make it more difficult for malicious actors to influence online discourse is allow social media users to opt-out of viewing ads posted by certain pages or accounts that they do not deem trustworthy or that do not operate from their province, state, region, or country during certain election or campaign advertising cycles.

## 2. News publishing: a departure from editorial standards

Before online platforms, participatory content in the news was limited to the presence of published letters to the editor or reader contributions in mainstream newspapers. This form of participation in the news is typically professionally edited, conforms to a strict set of guidelines regarding length, content, and style, and bears many other hallmarks of formal media. In the early years of the Internet, technologies enabling user-generated content consisted of decentralized general purpose tools such as email and HTML websites. With the development of digital platforms for specific use cases, users globally have been able to provide commentary on news topics, contribute to the

---

<sup>16</sup> “The Museum of Broadcast Communications - Encyclopedia of Television - Equal Time Rule.” Accessed February 19, 2018. <http://www.museum.tv/eotv/equaltimeru.htm>.

creation of news, and share in the dissemination of news via social media. This has contributed to the current news industry dynamic, which consists of dual participation by individuals as consumers and producers of news.

There are significant differences between the publishing and platform environments for news dissemination. Social media platforms, which operate as technology companies, have designed a system that replaces human editing with algorithms that elevate the “most engaging” content at the expense of less “viral” material. Social media platforms code the algorithms that apply these policies, thereby making editorial decisions about which content to prioritize and permit. These developments represent a departure from editorial standards in publishing news, which adhere to journalistic norms of objectivity and balance in broadcast and print media. By lowering the cost of entry to virtually anyone who can make a social media profile, social media platforms allow the amplification of “word of mouth” news, often allowing the speed of news transmission to take precedence over the truth, which has also undermined trust and credibility of news content available online by allowing the quick dissemination of fake news.

Methods introduced by social media companies to self-regulate their platforms against fake news have largely failed to address the full scope of the issue. While some platforms have experimented with machine-augmented fact-checking, they struggled to find automated solutions to tamp down on fake news because there is not always a clear line between true and false news online. Identifying the “truth” has been complicated for platforms, who do not want to become the arbiters of what is and is not true. Moreover, the standards for gauging the truth can be subjective; disputing a story can be debatable since what one source deems to be fake news may not be fake according to another source. In many situations, platforms refrain from taking action against partisan disinformation to avoid being seen as picking political sides. Importantly, fact-checking is often time-consuming, and fake posts can influence readers even if accompanied by a correction.<sup>17</sup> Platforms have also considered using user reports to mark false or inaccurate information. However, there exist concerns that such features could be used to “game the system” since other reporting tools have ended up being abused in a similar manner, with individual users finding their accounts suspended after organized campaigns resulted in hundreds of reports of “abusive” behaviour in a short space of time. Additionally, social media companies can be accused of being politically biased for removing certain posts after users report them as “false”. The scale of the platforms also means they can respond to fake posts and accounts often only in response to serious complaints once misinformation has already been spread.

## 2.1 Creating incentives that reward social media publishers based on content quality, accuracy or depth instead of only engagement metrics

Fake news proliferates based on the engagement-driven business model of social media platforms. For instance, when more people than usual click on a given news story on a platform like Facebook,

---

<sup>17</sup> Chris J Vargo, Lei Guo, Michelle A Amazeen, “The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016”, *New Media & Society*, June 15, 2017. <https://doi.org/10.1177/1461444817712086>

the company's software algorithms instantaneously spread and promote that story to many other users in the network, enabling articles to quickly "go viral" and making it harder to catch false news before it becomes ubiquitous. Since user attention is among the primary commodities traded on social media, and hyper partisan content garners greater attention, polarizing posts prompting emotive responses or fear mongering over appeals to reason tend to travel further. For instance, in the last months of the 2016 U.S. presidential election, the top 20 fake news pieces had greater engagement on Facebook than the top 20 stories from major news outlets.<sup>18</sup> Moreover, as an example, more engaging ads on Facebook have greater effective purchasing power in the platform's ad auctions. Despite modest changes to the Facebook News Feed algorithm to favor posts by friends over those by publishers<sup>19</sup> and reduce "engagement bait" in news feeds<sup>20</sup>, the most clicked and shared posts are given the largest and cheapest reach.

This incentive structure has unintended consequences for political and issue-based advertising. Social networks designed for virality and fast information flows also allow malicious actors to operate at scale and open new distribution channels for meddling in discourse or local affairs. There also exist monetary incentives to create more engaging ads that allow advertisers to reach larger audiences for the same amount of money. Apart from foreign adversaries who seek to undermine democratic elections, participants may also include non-malicious groups, such as the Macedonian teenagers who earned thousands of dollars by creating fake political news and news sites during the U.S. 2016 election campaigns.<sup>21</sup> While the mechanisms adopted by social media companies remain insufficiently robust to tackle the spread of fake news in this manner, further efforts need to be directed to ensure that more authentic and accurate content is given more widespread reach than are viral posts designed to misinform.

## Implementation

One solution could be rewarding verified news stories and posts by paying pages or publishers who consistently maintain journalistic standards more, which would encourage other publishers to adopt some of the same standards. This would incentivize publishing accurate, verified and high-quality news stories as opposed to retroactive fact-checking. Additionally, posts from more authoritative pages or publishers, or other indicators of quality content could be incorporated and used to prioritize quality in algorithmic content rankings to grant greater exposure to authentic content, making it harder to monetize the dissemination of fake news.

---

<sup>18</sup> Lee, Timothy B. "The Top 20 Fake News Stories Outperformed Real News at the End of the 2016 Campaign." Vox, November 16, 2016.

<https://www.vox.com/new-money/2016/11/16/13659840/facebook-fake-news-chart>.

<sup>19</sup> "Building a Better News Feed for You | Facebook Newsroom." Accessed February 19, 2018.

<https://newsroom.fb.com/news/2016/06/building-a-better-news-feed-for-you/>.

<sup>20</sup> "News Feed FYI: Fighting Engagement Bait on Facebook | Facebook Newsroom." Accessed March 30, 2018.

<https://newsroom.fb.com/news/2017/12/news-feed-fyi-fighting-engagement-bait-on-facebook/>.

<sup>21</sup> Wired, "Meet the Macedonian Teens Who Mastered Fake News and Corrupted the U.S. Election.", by Samantha Subramanian, February, 2017.

## 2.2 Offering tools for greater user control

Giving social media users more awareness and control over the filters used to hide or attain access to various types of ideas and news items would help to emulate real social networks where individuals, not tech companies, have more choice over the types and degrees of their exposure to different ideas and sources of news.

### Implementation

This could be achieved by offering tools to users for sorting their news feeds in different ways. Since individual users are the best judges of their own interests and preferences, by experimenting with and changing different filters based on their own priorities (of topics, communities, etc.), they would also be able to attain a greater understanding of why they see certain types of content in their news feeds and adjust their news filters according to their own preferences.

## 3. Bots, fake accounts and fake content: impersonation and disinformation at scale

Domestic and foreign actors on social media platforms can use bots or fake accounts to promote disinformation via “organic posts” without having to pay platforms for advertising. Platforms have so far relied on using analytical techniques including machine learning to detect and eliminate fake accounts operated by humans and bots.<sup>22</sup>

### 3.1 Labelling bot-administered accounts

Armies of bot-administered profiles pretending to be real human accounts computationally and automatically extend the ability to spread messages on social media platforms. Bots can be used to feign grassroots support for a policy, individual, or party when little such support exists (“astroturfing”) as well as spread norms and disinformation in a complex, contagious fashion to influence political discourse and manipulate public opinion online. When real users “like” posts shared by bots and share them on their own news feeds, small initial investments in advertising to boost posts can reach tens of millions of people.

While some research initiatives have begun more thorough study of the role of bots in disinformation and geopolitical interference, little is comprehensively known about the relative extent of bot activity on social media platforms. Additionally, while bot accounts on Twitter characteristically tweet frequently, retweet one another, and disseminate links to external content more often than human-operated accounts, studies have demonstrated the difficulty of detecting bot-operated accounts through any single criterion. Malicious bots, which have made up the majority of bot activity since 2013,<sup>23</sup> are meant to be unidentifiable by design and are notoriously

---

<sup>22</sup> Jen Weedon, William Nuland and Alex Stamos, “Information Operations and Facebook”, Facebook Security, April 27, 2017.

<sup>23</sup> “Freedom on the Net 2017: Manipulating Social Media to Undermine Democracy,” October 27, 2017. <https://freedomhouse.org/report/freedom-net/freedom-net-2017>.

difficult to counter from a technical standpoint. Social media users should be able to effectively recognize bot accounts in order to avoid risky interactions based on false assumptions.

#### Implementation

Given the difficulty users and tools face in detecting and stopping bots respectively, it is essential that bot-operated accounts and pages impersonating real humans are, at minimum, labelled to make them clearly distinguishable from human-administered profiles. News feed algorithms could also be amended to exclude or de-emphasize bot activity from lists of trending topics. Additionally, platforms should be required to have procedures in place for taking swift action when notified of influential bot accounts impersonating real persons or entities.

There is a need for further research in identifying bot activity on social media platforms and determining effective countermeasures. As an example, to counter the influence of social bots, Zhang et al. have proposed tracking each user's history of participating in spam distribution and suspending a user if their accumulated suspicious behaviors exceed a particular threshold.<sup>24</sup>

### 3.2 Verifying influential accounts and pages

The scale of the social media platforms means that inauthentic accounts and pages are often removed only in response to serious complaints, leaving the vast majority of user accounts and pages unverified. Despite this challenge, the issue of inauthentic accounts and pages can be an important driver for propagating fake news. Facebook in particular has identified that in the context of information operations, most false amplification on its platform is driven by coordinated people who are dedicated to operating inauthentic accounts, not automated processes such as bots.<sup>25</sup> Currently, Facebook's 'real name' policy mandates that account holders use their real names on the platform and users can submit government-issued IDs to verify that their profile indeed has their real name in order to avoid being banned, if reported.<sup>26</sup>

#### Implementation

Apart from offering account verification badges to all users who voluntarily submit their authentic government-issued IDs, influential accounts whose posts reach a substantial audience could be identified via a digital influence measurement scheme and subsequently offered options for verification. This can enable all account holders with significant influence and followings to have verified identities as opposed to monitoring for fake news content across large platforms. A simple example of such a measure would be to present verification options to all accounts that exceed a certain threshold for number of followers or whose posts have been shared and re-shared a substantial number of times. Enabling social media users to see which accounts have been offered verification options and which have been verified through various gradations of a check mark or

---

<sup>24</sup> Zhang, Jinxue, Rui Zhang, Yanchao Zhang, and Guanhua Yan. "The Rise of Social Botnets: Attacks and Countermeasures." *ArXiv:1603.02714 [Cs]*, March 8, 2016. <http://arxiv.org/abs/1603.02714>.

<sup>25</sup> Jen Weedon, William Nuland and Alex Stamos, "Information Operations and Facebook", Facebook Security, April 27, 2017.

<sup>26</sup> "Facebook Help Center." Accessed March 30, 2018. [https://www.facebook.com/help/159096464162185?helpref=faq\\_content](https://www.facebook.com/help/159096464162185?helpref=faq_content).

badge could help ensure a better perception of account authenticity. Similarly, if a group or page reaches a certain size or number of followers, the group or page administrator could be given options detailing how they can acquire a verified identity badge or mark. To an extent, this would ensure the authenticity of groups and pages that reach an influential following. At the same time, great care must be taken to ensure that users' fundamental rights or freedoms are not violated; for instance, accounts of activists or dissidents in closed societies should not be suppressed on the basis of an individual's inability to verify their identity if doing so would endanger their physical safety.

### 3.3 Authenticating fake video content

In addition to fake written content circulating on social media platforms, fake video content can be generated based on machine learning algorithms that can convert audio-clips into realistic, lip-synced videos of a person speaking the words.<sup>27</sup> The ability to control live or offline video in a seamless way using machine vision and image-based rendering has implications for tarnishing video evidence online.

#### Implementation

To detect the difference between spoofed and authentic video content featuring humans, some researchers have developed algorithms to determine whether the person on-screen has a human heartbeat by magnifying video frames and feeding them into a machine-learning program.<sup>28</sup> This built on work by an MIT team that pioneered Eulerian Video Magnification, a technique that can amplify and reveal small motions and run in real-time to indicate abnormalities indicative of spoofing such as widely varying heart rates.<sup>29</sup> Such techniques may be used by journalists or social media platforms to first detect and then promote authentic video content by emphasizing such authentic content in their respective news feed algorithms.

Another possible means of countering fake video and audio clips is to have publishers digitally sign and verify the content they produce and post on social media. While this would require widespread adoption to verify large amounts of video and audio content, social media platforms could also be required to tweak their news feed and results algorithms to favor and promote only verified and signed content.

---

<sup>27</sup> Supasorn Suwajanakorn, Steven M. Seitz, and Ira Kemelmacher-Shlizerman. "Synthesizing Obama: Learning Lip Sync from Audio". *ACM Trans. Graph.* 36, 4, Article 95, July 2017. DOI:

<http://dx.doi.org/10.1145/3072959.3073640>. Also see: Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., and Niessner, M. 2016. Face2Face: Real-time Face Capture and Reenactment of RGB Videos. In *Proc. CVPR 2016*. [https://web.stanford.edu/~zollhofer/papers/CVPR2016\\_Face2Face/paper.pdf](https://web.stanford.edu/~zollhofer/papers/CVPR2016_Face2Face/paper.pdf)

<sup>28</sup> "Real-Time Extraction of Biometric Data from Video." Accessed March 19, 2018.

[https://insights.sei.cmu.edu/sei\\_blog/2017/08/real-time-extraction-of-biometric-data-from-video-1.html](https://insights.sei.cmu.edu/sei_blog/2017/08/real-time-extraction-of-biometric-data-from-video-1.html).

<sup>29</sup> Wu, Hao-Yu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. "Eulerian Video Magnification for Revealing Subtle Changes in the World." *ACM Trans. Graph.* 31, no. 4 (July 2012): 65:1–65:8. <https://doi.org/10.1145/2185520.2185561>.

#### 4. Future research: next steps

While the aforementioned recommendations are important steps in the right direction, they provide just the starting point in tackling some of the complex socio-technical challenges posed by fake news. On their own, even a near perfect enforcement of the above recommendations is unlikely to solve the multi-faceted problem of fake news. More research must be undertaken to both understand the various aspects of the issues described above, and make additional design and regulatory decisions.

In particular, there is a need to further study issues concerning algorithmic transparency, AI accountability and privacy as it pertains to social media platforms. Social media platforms control individual user experiences, but continue to use algorithms operating as near black boxes that users do not sufficiently understand. Transparency regarding the way algorithms favor certain content over others would both encourage companies from engaging in questionable practices and help users better understand their news feeds, results and the ads targeted to them.

##### Implementation

Social media platforms should make the relevant aspects of their algorithms, data and practices available for further research to address gaps in knowledge about the actual practices of targeting and tailoring advertising to users, filtering information and news content, and the activity of bots and inauthentic accounts on their platforms among other research areas. Greater government and third-party funding from unbiased sources into these issues could help address various concerns about algorithmic transparency, determine appropriate standards for auditing algorithms and platforms, and establish industry-wide best practices to ensure that the algorithms and practices used do not contribute to the prevalence of fake news on social media platforms.